Draft Methods Guide for Comparative Effectiveness Reviews

Prioritization and Selection of Harms for Inclusion in Systematic Reviews

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
5600 Fishers Lane
Rockville, MD 20857
www.ahrq.gov

Contract No.

Prepared by:

Investigators:

AHRQ Publication No. xx-EHCxxx
<Month Year>

This report is based on research conducted by the Agency for Healthcare Research and Quality (AHRQ) Evidence-based Practice Centers' Methods Workgroup. The findings and conclusions in this document are those of the authors, who are responsible for its contents; the findings and conclusions do not necessarily represent the views of AHRQ. Therefore, no statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

This research was funded through contracts from the Agency for Healthcare Research and Quality.

The information in this report is intended to help health care decisionmakers—patients and clinicians, health system leaders, and policy makers, among others—make well-informed decisions and thereby improve the quality of health care services. This report is not intended to be a substitute for the application of clinical judgment. Anyone who makes decisions concerning the provision of clinical care should consider this report in the same way as any medical reference and in conjunction with all other pertinent information (i.e., in the context of available resources and circumstances presented by individual patients).

None of the investigators have any affiliations or financial involvement that conflicts with the material presented in this report.

Persons using assistive technology may not be able to fully access information in this report. For assistance, contact EffectiveHealthCare@ahrq.hhs.gov

Suggested Citation: Pending

Preface

The Agency for Healthcare Research and Quality (AHRQ), through its Evidence-based Practice Centers (EPCs), sponsors the development of evidence reports and technology assessments to assist public- and private-sector organizations in their efforts to improve the quality of health care in the United States. The reports and assessments provide organizations with comprehensive, science-based information on common, costly medical conditions and new health care technologies and strategies. The EPCs systematically review the relevant scientific literature on topics assigned to them by AHRQ and conduct additional analyses when appropriate prior to developing their reports and assessments.

To improve the scientific rigor of these evidence reports, AHRQ supports empiric research by the EPCs to help understand or improve complex methodologic issues in systematic reviews. These methods research projects are intended to contribute to the research base in and be used to improve the science of systematic reviews. They are not intended to be guidance to the EPC program, although they may be considered by EPCs along with other scientific research when determining EPC program methods guidance.

AHRQ expects that the EPC evidence reports and technology assessments will inform individual health plans, providers, and purchasers and the health care system as a whole by providing important information to help improve health care quality. The reports undergo peer review prior to their release as a final report.

We welcome comments on this Methods Research Project. They may be sent by mail to the Task Order Officer named below at: Agency for Healthcare Research and Quality, 5600 Fishers Lane

Rockville, MD 20857, or by e-mail to epc@ahrq.hhs.gov.

Andrew B. Bindman, M.D.
Director
Agency for Healthcare Research and Quality

Arlene S. Bierman, M.D., M.S.
Director
Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality

Stephanie Chang, M.D., M.P.H.
Director
Evidence-based Practice Center Program
Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality

Lionel Bañez, M.D.
Task Order Officer
Evidence-based Practice Center Program
Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality

Contents

Introduction

The Agency for Healthcare Research and Quality's (AHRQ) Evidence-based Practice Center (EPC) Program has long recognized the need for systematic reviews of medical interventions to provide balanced assessments that include evaluation of harms as well as benefits. However, synthesizing evidence on harms poses unique challenges. The assessment and reporting of harms is often suboptimal,[1-3] studies are often too short to evaluate important long-term harms and have inadequate statistical power to evaluate serious but uncommon harms,[5, 6] patients enrolled in research studies are frequently at lower risk for harms than those encountered in clinical practice,[7] potentially resulting in underestimation of harms, and important data on harms may be unpublished or selectively reported.[9-11]

In 2005, AHRQ funded a series of white papers on challenges in evidence synthesis that included an article on evaluation of harms.[5] It highlighted unique challenges in finding and selecting data on harms, rating the quality of harms reporting, and synthesizing and displaying from studies reporting harms. Subsequently, recommendations for synthesizing evidence on harms were further developed by a Methods Workgroup of EPC investigators convened by AHRQ; these recommendations were codified in 2010 as a chapter in the AHRQ EPC Program Methods Guide.[13] Issues addressed by the Workgroup included the need to consider a broad range of data sources to evaluate harms, including observational studies as well as randomized controlled trials and unpublished as well as published data; the importance of using consistent and precise terminology on harms; the need to evaluate the quality of harms assessment and reporting distinctly from rigor for assessing benefits; and challenges in synthesis, including evaluation of rare events, use of indirect comparisons, and pooling methods.

In 2015, AHRQ convened an EPC Methods Workgroup to update or expand upon prior guidance for assessing harms. Following initial deliberations, the Workgroup elected to focus on updating guidance on prioritization and selection of harms in systematic reviews. Although the 2010 harms chapter recommended that EPC reviews "always assess harms that are important to clinicians and patients," the Workgroup found that it lacked more specific recommendations regarding how to select the harms to be included in an EPC review, and noted that selection and prioritization of harms in EPC reviews poses important challenges. EPC reviews frequently address many interventions, which could result in many (e.g., dozens) of potential harms to review. Unlike benefits, which are often similar across interventions used to treat a given condition (e.g., medications, non-pharmacological therapies, and surgery for low back pain are all aimed at improving pain and function), different interventions given for the same condition are frequently associated with a large number of disparate harms. For example, medications for low back pain are typically associated with a set of harms distinct from those associated with surgery, and different medications are each associated with unique harms. Other issues include whether to assess nonspecific harms (e.g., "serious harms" or "withdrawal due to adverse events"), which might facilitate comparisons between interventions with dissimilar harms, and how to address unanticipated harms that are encountered during the review process. Workgroup members noted that including all potential harms in these cases is not feasible, and can make it difficult for users of EPC reports to reach bottom-line conclusions regarding harms or determine the balance of benefits to harms. Workgroup members noted that clearer methods would be helpful for supporting the decisions made regarding selection of harms and help focus EPC reports on the outcomes of greatest importance, potentially increasing their usability. The

immediate intended audience of this guidance is the EPC program, though we hope it may be useful to all systematic reviewers and those who commission or use systematic reviews

Methods

Approach

We assembled a workgroup of 12 methodologists from AHRQ, the EPC program, and the Scientific Resource Center (SRC) to develop recommendations on selection and prioritization of harms, building on previous work by a prior EPC Workgroup. The project was led by an investigator at the Pacific Northwest EPC. Members participated in twice monthly teleconference calls over the span of 11 months and sought information through a literature scan, a review of EPC reports, and interviews with Key Informants, in order to inform the development of consensus recommendations.

Literature search and review

The Scientific Resource Center (SRC) provides support for the AHRQ EPC Program for the advancement of scientific methods, strategic planning, peer review, topic nomination and education. As part of this work, the SRC curates a bibliographic database of nearly 10,000 citations on the methodology of systematic reviews and comparative effectiveness research, dating back to the 1950s.[15] On November 10, 2015, the SRC librarian performed a keyword search and a descriptor search for "Harms/Adverse Events" in the SRC Methods Library database (n=357). The citations were filtered from a publication date of 2007 on, so as to only include more recent articles, including studies published since the prior harms chapter in the AHRQ EPC Program Methods Guide (n=257).[13] Two members of the workgroup then conducted a dual review of the citations, seeking articles that could provide guidance on the methods for selecting and prioritizing harms for inclusion in systematic review, or that reported empiric research in that area. We also searched for articles that could provide guidance on the methods for selecting and prioritizing harms for inclusion in systematic review, or that reported empiric research in that area (beneficial or harmful) for information relevant to prioritization and selection of harms. Because we were seeking literature that could inform discussions and context and anticipated that empiric research would be sparse, we did not apply strict eligibility criteria.

Review of EPC reports

An SRC investigator reviewed 18 EPC reports to determine year of publication, key questions related to harms and the harms that were assessed, methods used to prioritize or select harms, the data sources used to identify evidence on harms, and main findings regarding rates of harms. We categorized the harms assessed as "specific" or "non-specific harms" included composite measures of various harms such as presence of any harm, serious harms, withdrawal or discontinuation due to adverse events, or similar. We selected EPC reports published since the year 2014 so that the sample represents recent methods. We did not review a random sample of EPC reports, but instead selected reports to provide a representative sample from various EPCs and diversity in terms of the types of interventions evaluated and compared. Information from the EPC reports was abstracted into an Excel spreadsheet, which was provided to Workgroup members to inform discussions.

Key Informant interviews

The SRC compiled a list of methodologist key informants (KIs) with experience in conducting, commissioning, or using systematic reviews for interviewing. We then ranked first priority and second priority key informants based on organization and experience. On February 19, 2016 the SRC sent email invites to (n=14) first priority KIs, providing them with background on the project and the purpose of the requested interview (Appendix A).

The workgroup lead, Roger Chou, with input from the group, compiled a list of 12 targeted questions with the aim of informing our discussion on the selection and prioritization of harms in systematic reviews (Appendix B). The interview guide was sent to KIs prior to their interview and covered these general topics:

- Use of published guidance for prioritization and selection of harms
- Criteria for prioritizing harms
- Using input from stakeholders to guide prioritization and selection of harms
- Using literature and other data sources to guide prioritization and selection of harms
- Thresholds for maximum number of harms to be reviewed
- Inclusion of non-specific harms
- Methods for addressing unanticipated harms
- Reporting of methods for prioritizing and selection of harms.

KIs also completed a conflict of interest form prior to their participation; none were determined to have conflicts that precluded their participation. Over the span of 5 months (March 2016 – July 2016), workgroup members conducted 5 telephone interviews with 6 KIs, lasting 60 minutes each. Each interview was recorded, transcribed, and sent to its respective KI as an opportunity for further elaboration, clarification, and corrections. We identified common themes in the responses to interview questions across the KIs, as well as areas in which responses differed. Notes from each interview and a document summarizing the themes were then presented to the Workgroup for further discussion (Appendix C).

Development of recommendations

All Workgroup members reviewed the results of the literature scan, review of EPC reports, and notes from the KI interviews and summarized themes, which were discussed on regularly scheduled conference calls. A draft set of recommendations was developed and distributed to the Workgroup for further discussion and feedback. Given the lack of strong empiric evidence to guide recommendations in this area, the Workgroup sought to reach consensus on all recommendations.

Results

Literature search

After screening 257 citations from the SRC methods research database and reviewing 108 full-text articles, we identified no empiric research on the utility or validity of different methods to prioritize harms in systemic reviews to inform our discussions. Although several articles provided general guidance on assessment of harms or on prioritization of outcomes for systematic reviews, none provided recommendations specifically on selection and prioritization of harms, other than the prior EPC methods work.[13] The harms chapter of the EPC Methods Guide recommended that reviewers assess the harms that are important to decision makers and users of the intervention under consideration; it noted that high-priority harms are the most serious adverse events and may include common adverse events or other adverse events important to clinicians and patients. It suggested that systematic review authors use prior reviews, safety reports from the US Food and Drug Administration, postmarketing surveillance databases, and input from technical experts and patients to identify and prioritize harms to be evaluated. We also reviewed articles providing general guidance on conduct of systematic reviews and synthesis of evidence, but found little guidance on selection and prioritization of harms. For example, regarding selection of harms, the Cochrane Handbook notes the harms selected for a review depend on the study question and the therapeutic or preventive context, and that the reviewer could opt for a narrow focus (e.g., one or two known or a few of the most serious adverse effects that are of special concern to patients and health professionals) or a broad focus (e.g., the 5 to 10 most frequent adverse effects, all adverse effects that either the patient or clinician consider to be serious, or organized by category [e.g., diagnosed by lab results or patient-reported symptoms]).[16] The GRADE working group recommends that guideline developers prioritize outcomes (both beneficial and harmful), which can be done through solicitation of panel member and stakeholder input and using a 1-9 numerical rating system.[18] The outcomes rated highest priority are the ones that the guideline development group will focus on in assessing the balance of benefits to harms and informing recommendations. The GRADE working group recommends that summary of findings tables focus on no more than 7 of the most patient-important outcomes (including both beneficial and harmful outcomes), in order to avoid overwhelming the reader, while providing information on the most critical outcomes.[20]

Review of EPC reports

We reviewed 18 EPC reports that addressed a range of intervention types (e.g., medical, surgical, diagnostic testing, informatics, behavioral therapy) and conditions (e.g., cancer, musculoskeletal, surgical, psychological, lipids, obstetric, neurological, otolaryngologic). All of the reports had key questions related to harms. No EPC report described the method used to select or prioritize harms, though most reported on serious and common harms, implying that severity and frequency guided decisions regarding which harms to include. Few EPC reports described results for nonspecific harms such as “any adverse event,” “withdrawal due to adverse event,” or “any serious adverse event”; rather, the reports generally focused on specific adverse events, sometimes categorizing their severity. No EPC report described using a formal prioritization process or the sources used to inform decisions regarding which harms to include.

Synthesis of KI interviews

Use of published guidance for prioritization and selection of harms

None of the KIs reported using published guidance for prioritization and selection of harms. Although the KIs were generally aware of GRADE methods for prioritization of outcomes, none reported using GRADE methods to prioritize harms, and few had experience applying a formal GRADE prioritization process to selection of outcomes in general. Several KIs were aware of published guidance from the Cochrane Collaboration and AHRQ on assessing harms in systematic reviews, but were not aware of specific guidance on prioritization and selection of harms.

Criteria for prioritizing harms

All of the KIs noted that systematic reviews should prioritize harms that are of most importance to decision makers. They noted that these typically include serious harms as well as less-serious but common harms. The KIs noted that severity of harms is often poorly or inconsistently defined, which makes determining whether a harm is "serious" a challenge, though they also noted that there are published definitions for categorizing seriousness of harms (e.g., the Food and Drug Administration criteria for reportable "serious" adverse events).[22] The KIs noted that the quality or quantity of evidence should not be an important factor in selection and prioritization of harms; rather, they emphasized the need to select and prioritize harms that are important to decision makers, regardless of the evidence and evidence sources available.

Using input from stakeholders to guide prioritization and selection of harms

All KIs described obtaining input from clinical and content experts to inform decisions regarding prioritization and selection of harms. Although most KIs described a relatively informal process (e.g., soliciting general feedback from a panel of stakeholders on outcomes to be addressed in a conference call or electronically), others described a more formal process in which KIs were asked to rank or rate outcomes related to harms. Several KIs also described using input from patient stakeholders. In some cases, patients participated in a larger group with clinical or content experts and in others, patients provided input separately. The KIs noted that a challenge in engaging stakeholders to prioritize harms is that clinicians and patients could prioritize harms differently.[23] One KI described a process in which patients were asked to rank/prioritize outcomes and described challenges in interpreting or using the findings, such as patients rating all outcomes as similarly high priority (e.g., mortality and an intermediate laboratory outcome both prioritized similarly) or patients having difficulty understanding the systematic review process or the scientific issues. Another KI noted that her organization had convened a group of patient stakeholders who are to receive training in systematic review methods who would be asked to provide input for multiple reviews on an ongoing basis.

Using literature and other data sources to guide prioritization and selection of harms

All of the KIs noted that a broad range of data sources should be utilized to inform decisions regarding which harms to consider for inclusion. Suggested data sources included randomized controlled trials, observational studies (including pharmacoepidemiological studies performed on large databases), and information from regulatory agencies and other groups that collect postmarketing information and case reports on adverse events (e.g., the FDA's MedWatch program).[25] The KIs noted that reviewers should

not rely solely on randomized trials since they are often underpowered to detect uncommon harms, too short to evaluate long-term harms, and often enroll "ideal" populations at low risk for harms. The KIs noted that case reports may identify potentially serious harms that are very uncommon; however, they also noted that it can be difficult to determine causality from such studies.

Thresholds for maximum number of harms to be reviewed

The KIs were generally aware that the GRADE Working Group has suggested a maximum threshold of outcomes to include when summarizing the evidence. However, they felt that it was difficult to apply a maximum threshold for harms to be included in EPC reports, given the large number of interventions and comparisons that are often being evaluated. In addition, the KIs noted that even for a single comparison, limiting to a maximum of 7 beneficial and harmful outcomes as recommended by the GRADE Working Group could result in only 3 or 4 harms, which they felt was fewer than necessary to adequately assess the harms of most interventions. Nonetheless, the KIs agreed that it is important to focus the EPC reports on the most critical harms, without applying a specific threshold for the maximum number to be evaluated, in order to help make the reports more usable.

Inclusion of non-specific harms

The KIs generally felt that inclusion of non-specific harms could be helpful in facilitating comparisons between interventions with dissimilar harms. They noted that non-specific harms could be considered composite outcomes since they consist of a variety of different harms, and must be interpreted with caution. They suggested that if non-specific harms are included, it would generally be more useful to focus on more severe harms, as indicated by "any serious harm" or "withdrawals due to adverse events," rather than non-specific harms that include less serious events (e.g., "any adverse event"), which may be more difficult to interpret. The KIs also noted that non-specific harms should be interpreted in conjunction with data on specific harms.

Methods for addressing unanticipated harms

The KIs noted that during the review process, reviewers may encounter or become aware of potentially relevant harms not considered in the prioritization process ("unanticipated harms"). This discovery could be due to the publication of new data or analyses, or patterns/data that the reviewers observe in the course of conducting the review. The KIs suggested that reviewers should be open to including information on unanticipated harms not identified during the protocol development phase, and be prepared to modify the study protocol to note their inclusion. However, they also expressed the belief that data and analyses regarding unanticipated harms should generally be considered hypothesis generating and presented as such. KIs indicated that for unanticipated harms for which data appeared compelling, reviewers should consider proposing future research to clarify potential associations. The reviewers noted examples in which unanticipated harms ended up not being clearly confirmed in prospective studies (e.g., increased myocardial infarction with thiazolinediones[27, 29] as well as examples in which unanticipated harms have been confirmed in subsequent analyses (e.g., increased myocardial infarction risk with cyclo-oxygenase-2-selective non-steroidal anti-inflammatory agents).[31]

Reporting of methods for prioritizing and selection of harms

The KIs agreed that systematic reviews typically do not report methods used to select or prioritize harms; this was consistent with our review of EPC reports. The KIs felt that it would be helpful for systematic

reviews to report any prioritization methods used, including how stakeholders were engaged, criteria used to determine which harms were included (e.g., seriousness, frequency), and the criteria used to select included harms. The KIs noted that providing this methodologic information would help readers better understand the basis for prioritization decisions.

Recommendations

Prioritization of harms

1) Include harms that are of most importance to decision-makers.

The workgroup recommends that EPC reviews include the harms judged to be of most importance to decision-makers. Typically, these will be serious harms as well as less serious but common and/or bothersome harms. Using the FDA definition, serious harms are those that result in death, are life-threatening, result in hospitalization or prolongation of an existing hospitalization, result in persistent or significant incapacity or ability to perform normal life functions, or are congenital anomalies or birth defects.[33] Other harms may also be considered serious when judged to jeopardize the patient or subject and may require medical or surgical intervention to prevent one of the outcomes listed above. From the perspective of a decision-making framework, a harm may be considered "important" if the probability of that harm occurring likely to expected benefits would impact recommendations about the use of the intervention. An exception to routinely including all harms of most importance is reviews that focus on a specific, pre-defined harm or harms (this is not typical for EPC reviews); in these cases the scope of the review should be clearly explained.

2) Use a prioritization process to help narrow the number of harms included in a review.

Recognizing that it will often not be feasible to include all potential harms in an EPC review, the workgroup recommends that EPCs utilize a process to prioritize the harms of most importance to be reviewed. Generally speaking, the harms prioritized in this process will be included in summary of evidence tables, along with prioritized benefits. Using a prioritization process will help strengthen the rationale for the harms that are selected for review and provide a basis for the selection decisions that are made.

3) The prioritization may be informal (e.g., input or informal interviews with experts in the field, patients, and other stakeholders, literature scan, review of information from regulatory agencies) or more formal (e.g., Delphi or GRADE-like scoring process).

Although the workgroup suggests that EPC utilize a prioritization process, it found insufficient evidence to recommend a specific prioritization method. The workgroup suggests that EPCs obtain input from stakeholders, including clinical/technical experts, policymakers, and patients; perform a literature scan; and review information from regulatory agencies to inform the prioritization process. Although more formal prioritization methods may be useful (e.g., formal consensus process or use of a GRADE-like scoring/prioritization method), the workgroup concluded that it is unclear whether using such methods results in more appropriately selected/prioritized harms than less formal processes. As noted by some KIs, incorporating stakeholder input in a more formal process also could be a challenge, e.g. when

different stakeholders prioritize harms differently or when very serious clinical outcomes and minor harms (e.g., laboratory based intermediate outcomes) are prioritized similarly. In addition, utilizing such processes impact the time and resources required to conduct the review. Therefore, until more data are available on the effects of using more formal prioritization processes on the usefulness and credibility of systematic reviews, the workgroup concluded that a recommendation for their formal use was not warranted.

4) The method used to prioritize harms should be concordant with methods used to select outcomes related to benefit.

The workgroup recommends that the methods used to prioritize and select harms be concordant with the methods used to prioritize and select beneficial outcomes, given that the principles underlying the prioritization of outcomes are similar, whether they are to measure beneficial or harmful effects. The workgroup acknowledged that prioritization of beneficial outcomes is often more straightforward than for harmful outcomes since the expected benefits for different interventions administered for the same condition are often similar. However, as for harmful outcomes, there may be many potential beneficial outcomes to consider. It may be difficult to distinguish harms from failed treatments (e.g., myocardial infarction in patient on statin therapy); whether an event is classified as a benefit or harm may depend on the intended effect of the treatment and the perspective of the decision maker.

Type of harms to include

5) Routinely include serious harms or less serious but frequent or bothersome harms, or describe why they aren't included.

As noted above, the workgroup recommends that EPC reviews routinely include serious harms or less serious but frequent or bothersome harms. In some cases, EPC reviews may not include all such harms. This could be because the harms are well-established and do not require another review; the intervention is not thought to be associated with major harms (e.g., eyeglasses for decreased visual acuity, hearing aids for hearing loss, ultrasound for musculoskeletal conditions); or the review is focused on a particular harm or harms. When applicable, such circumstances should be explained. In general, intermediate outcomes (e.g., changes in laboratory values or physiological parameters) are considered lower priority than patient-centered health outcomes (e.g., mortality or outcomes related to morbidity, qualify of life, or function). EPC reviews may consider inclusion of intermediate outcomes related to harms when data on associated clinical outcomes is sparse and the association between intermediate outcomes and clinical harms is well established (e.g., severe anemia or neutropenia).

6) Non-specific adverse events may help facilitate comparisons across interventions; routinely consider including "serious adverse events" or "withdrawal due to adverse events," particularly when evaluating head-to-head comparisons.

The workgroup recommends that EPC reviewers consider including non-specific adverse events, which may help facilitate head-to-head comparisons, particularly for interventions associated with dissimilar harms. The workgroup suggests that EPC reviews focus on indicators of more severe harms (e.g., "serious" adverse events or "withdrawal due to adverse events"), given the non-specific adverse events are a composite outcome and it is more difficult to interpret the clinical meaningfulness of less severe harms. The workgroup recommends that EPC reviews not focus solely on non-specific adverse events;

rather, non-specific harms should be interpreted in conjunction with data on the specific harms that comprise these composite outcomes.

7) For reviews that involve effects of diagnostic tests, consider inclusion of over-diagnosis and overtreatment, as well as other harms related to diagnostic testing (e.g., false-positives and –negatives and effects thereof, labeling, and others).

Although intermediate outcomes are generally considered lower priority than clinical outcomes, for certain interventions (e.g., those addressing diagnostic testing), the workgroup suggests that EPC reviews consider inclusion of intermediate measures of harm such as overdiagnosis or overtreatment as a result of testing. Such outcomes may help identify important negative downstream effects of testing that are otherwise difficult to capture. The workgroup acknowledges challenges in measuring these outcomes, and variability in the methods used. [34, 36] Other harms associated with diagnostic tests include false-positives and negatives and the consequences of such findings, labeling, and others.

Number of harms to include

8) A reasonable rule of thumb is to limit to 5-10 prioritized harms for each comparison dyad, though there is no preset threshold for the number of harms selected for a review. Including more harms may make it difficult for users to process/interpret the findings.

Given the large number of interventions and comparisons that may be included in an EPC review, the number of potential harms to be reviewed may be overwhelming and difficult to process for users of the reviews, in order to assess trade-offs between potential benefits and harms. Therefore, the workgroup recommends that EPC reviews limit the number of prioritized harms to be reviewed. The workgroup felt that using the suggested GRADE maximum threshold of 7 beneficial and harmful outcomes would frequently result in exclusion of potential important harms. Instead, it suggests that EPC reviewers utilize an approach that is based on the number of comparisons. For each comparison dyad, the workgroup suggests that the EPC aim for 5-10 prioritized harms (including specific as well as non-specific harms). Across dyads, to the extent possible the workgroup suggests that EPCs identify common prioritized harms, in order to limit the total number of harms to be assessed. For reviews in which there are many comparisons and potential harms, the workgroup suggests that reviewers aim for a number of harms selected for each comparison dyad on the lower end of the range.

Unanticipated harms

9) Be prepared to add harms that are not specified in the protocol or considered in the prioritization process; findings for unanticipated harms will often be considered hypothesis generating.

The workgroup recommends that EPC reviews be prepared to incorporate unanticipated harms into the review. Unanticipated harms may be identified during the course of data analysis of included studies, or via outside sources (e.g., new published study, regulatory agency action). Because unanticipated harms are not prespecified, their addition should be recorded as a protocol modification. In addition, the workgroup suggests that EPC reviews clearly indicate findings related to unanticipated harms. EPCs should interpret findings related to unanticipated harms in the context of other information, including the plausibility of biological mechanisms of action, pharmacokinetic data, and other published data on the

harm that may have previously been overlooked or unidentified.[38] In many cases, findings for unanticipated harms will be considered hypothesis-generating.

Reporting methods used to select harms

10) Report the methods used to prioritize harms, differentiate serious from frequent but less serious harms, and indicate interventions for which serious harms are not believed to be an issue and why.

The workgroup recommends that EPC reviews report methods used to prioritize harms, including the composition of stakeholder groups providing input, literature scan methods, and other data sources. In addition, EPC reviews should describe the prioritization process, whether informal or more formal. EPC reviews should differentiate which harms are considered serious and those considered less serious but of high frequency or most bothersome. In situations in which serious harms are not included, EPC reports should provide the reason (e.g., the intervention is not believed to be associated with serious harms, serious harms have already been established, the review is scoped to focus on a specific harm or harms).

Discussion

Selection or prioritization of harms in EPC reviews is an important challenge that has not previously been addressed in depth in the EPC Methods Guide. Although EPC reviews seek to be comprehensive and provide balanced assessments of benefits and harms, inclusion of multiple interventions and comparisons often results in many potential harms to be reviewed, which could be overwhelming to users. A review of EPC reports indicate that they provide little or no information regarding how harms were selected. A scan of the literature found little guidance on selection and prioritization of harms in systematic reviews. This article provides guidance developed by a workgroup of EPC methodologists on selection and prioritization of harms. Key recommendations include: routinely focusing on serious as well as less serious but frequent or bothersome harms; routinely engaging stakeholders and using literature scans and other data sources to identify harms of importance; using a prioritization process (whether formal or less formal) to inform prioritization decisions; and describing the methods used to select and prioritize harms. For future priorities in developing guidance on assessing harms in EPC reviews, the workgroup identified methods for assessing the quality of harms reporting and determining which types of studies to use to evaluate harms as high priority topics. The workgroup recognizes that the data supporting the recommendations in this article are sparse and that follow-up to assess the impact of the recommendations on reporting, usefulness/usability of reports, and appropriateness of prioritization decisions is needed.

References

1. Haidich AB, Birtsou C, Dardavessis T, et al. The quality of safety reporting in trials is still suboptimal: survey of major general medical journals. Journal of clinical epidemiology. 2011;64(2):124-35.
2. Garritty CM. Updating systematic reviews: the policies and practices of health care organizations involved in evidence synthesis. Ottawa, CA: Library and Archives Canada = Bibliothèque et Archives Canada; 2009.
3. Hodkinson A, Kirkham JJ, Tudur-Smith C, et al. Reporting of harms data in RCTs: a systematic review of empirical assessments against the CONSORT harms extension. BMJ open. 2013;3(9):e003436-2013-.
4. Ahmed I, Sutton AJ, Riley RD. Assessment of publication bias, selection bias, and unavailable data in meta-analyses using individual participant data: a database survey. BMJ. 2012;344:d7762.
5. Chou R, Helfand M. Challenges in systematic reviews that assess treatment harms. Annals of Internal Medicine. 2005;142(12 Pt 2):1090-9.
6. Bhaumik DK, Amatya A, Normand SL, et al. Meta-Analysis of Rare Binary Adverse Event Data. Journal of the American Statistical Association. 2012;107(498):555-67.
7. Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?". The Lancet. 2005;365(9453):82-93.
8. Garritty CM. Chapter 4: A framework for updating systematic reviews. Ottawa, ON: Dalla Lana School of Public Health , University of Toronto; 2009.
9. Hodkinson A, Gamble C, Smith CT. Reporting of harms outcomes: a comparison of journal publications with unpublished clinical study reports of orlistat trials. Trials. 2016;17(207).
10. Gartlehner G, Dobrescu A, Evans TS, et al. Average effect estimates remain similar as evidence evolves from single trials to high-quality bodies of evidence: a meta-epidemiologic study. Journal of Clinical Epidemiology. 2016;69(1):16-22.
11. Saini P, Loke YK, Gamble C, et al. Selective reporting bias of harm outcomes within studies: findings from a cohort of systematic reviews. BMJ. 2014;349(Journal Article):g6501.
12. Ding Y, Fu H. Bayesian indirect and mixed treatment comparisons across longitudinal time points. Statistics in Medicine. 2013;32(15):2613-28.
13. Chou R, Aronson N, Atkins D, et al. AHRQ series paper 4: assessing harms when comparing medical interventions: AHRQ and the effective health-care program. Journal of clinical epidemiology. 2010;63(5):502-12.
14. Berry D, Wathen JK, Newell M. Bayesian model averaging in meta-analysis: Authors' response. Clinical Trials. 2009;6(1):50-1.
15. AHRQ Secure Site Methods Article Alert - EHC Scientific Resource Center EPC-SRC.org: EPC-SRC.org; 2016. http://epc-src.org/methodsLibrary/artAlert.cfm. Accessed on 2016.
16. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions. The Cochrane Collaboration. 2011.
17. Gasparrini AAB, Kenward MG. Multivariate meta-analysis for non-linear and other multi-parameter associations. Statistics in medicine. 2012.
18. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 2. Framing the question and deciding on important outcomes.

Journal of clinical epidemiology. 2011;64(4):395-400.
19. Gasparrini A, Armstrong B. Reducing and meta-analysing estimates from distributed lag non-linear models. BMC Medical Research Methodology. 2013;13:1-2288-13-1.
20. Guyatt GH, Oxman AD, Santesso N, et al. GRADE guidelines: 12. Preparing Summary of Findings tables—binary outcomes Journal of clinical epidemiology. 2013;66(2):158-72.
21. Gasparrini A, Armstrong B. Multivariate meta-analysis: a method to summarize non-linear associations. Statistics in medicine. 2011;30(20):2504-6; discussion 9-10.
22. CFR - Code of Federal Regulations Title 21 Sec. 312.32. https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfCFR/CFRSearch.cfm?fr=312.32: U.S. Food and Drug Administration; 2016. https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfCFR/CFRSearch.cfm?fr=312.32.
23. Montgomery A, Fahey T. How do patients' treatment preferences compare with those of clinicians? Quality Health Care. 2001;10(1).
24. Gasparrini A, Armstrong B, Kenward MG. Multivariate meta-analysis for non-linear and other multi-parameter associations. Statistics in Medicine. 2012;31(29):3821-39.
25. Craigle V. MedWatch: The FDA Safety Information and Adverse Event Reporting Program. Journal of the Medical Library Association. 2007;95(2):224-5.
26. Gerger H, Munder T, Gemperli A, et al. Integrating fragmented evidence by network meta-analysis: relative effectiveness of psychological interventions for adults with post-traumatic stress disorder. Psychological Medicine. 2014;44(15):3151-64.
27. Dormandy J.A. CB, Eckland D.J., Erdmann E., et al. Secondary prevention of macrovascular events in patients with type 2 diabetes in the PROactive Study (PROspective pioglitAzone Clinical Trial In macroVascular Events): a randomised controlled trial. The Lancet. 2005;366(9493):1279-89.
28. Ghania AFSA. Some Contributions to Meta Analysis Quasi-Empirical Bayes Method of Estimation. Ottowa, Ontario, Canada: Carlton University; 2011.
29. Home P.D., Pocock S.J., Beck-Nielsen H., et al. Rosiglitazone evaluated for cardiovascular outcomes in oral agent combination therapy for type 2 diabetes (RECORD): a multicentre, randomised, open-label trial. The Lancet. 2009;373(9681):2125-35.
30. Ghidey W, Stijnen T, van Houwelingen HC. Modelling the effect of baseline risk in meta-analysis: A review from the perspective of errors-in-variables regression. Statistical Methods in Medical Research. 2013;22(3):307-23.
31. Trelle S, Reichenbach S, Wandel S, et al. Cardiovascular safety of non-steroidal anti-inflammatory drugs: network meta-analysis. BMJ. 2011;342(7086).
32. Ghosh D, Taylor JM, Sargent DJ. Meta-analysis for surrogacy: accelerated failure time models and semicompeting risks modeling. Biometrics. 2012;68(1):226-32.
33. What is a Serious Adverse Event? http://www.fda.gov/Safety/MedWatch/HowToReport/ucm053087.htm: U.S. Food and Drug Administration; 02/01/2016 2016.
34. Loeb S., Bjurlin M.A., Nicholson J., et al. Overdiagnosis and overtreatment of prostate cancer. European urology. 2014;65(6):1046-55.
35. Gibbons RD, Amatya A. Statistical Methods for Drug Safety. Chapman and Hall/CRC: Boca Raton, FL; 2015.
36. Jha S. An Overview of Methodologies in Detecting Overdiagnosis. Academic Radiology. 2015;22(8):1046-7.
37. Gil-Herrera E, Tsalatsanis A, Kumar A, et al. Identifying homogenous subgroups for

individual patient meta-analysis based on Rough Set Theory. Conference proceedings : ...Annual International Conference of the IEEE Engineering in Medicine and Biology Society.IEEE Engineering in Medicine and Biology Society.Annual Conference. 2014;2014:3434-7.

38. Avorn J. Evaluating drug effects in the post-vioxx world there must be a better way. Circulation. 2006;113(18):2173-6.